The information-theoretic meaning of Gagliardo–Nirenberg type inequalities

Giuseppe Toscani *

Dedicated to the memory of Emilio Gagliardo

Abstract. Gagliardo–Nirenberg inequalities are interpolation inequalities which were proved independently by Gagliardo and Nirenberg in the late fifties. In recent years, their connections with theoretic aspects of information theory and nonlinear diffusion equations allowed to obtain some of them in optimal form, by recovering both the sharp constants and the explicit form of the optimizers. In this note, at the light of these recent researches, we review the main connections between Shannon-type entropies, diffusion equations and a class of these inequalities.

Keywords. Heat equation, nonlinear diffusions, Shannon entropy, Rényi entropy, Fisher-type informations, Gagliardo–Nirenberg type inequalities.

1 Introduction

Gagliardo–Nirenberg inequalities, proved independently by Gagliardo in [27] and Nirenberg [37] in the late fifties, are interpolation inequalities of the form

$$\|u\|_{L^p(\mathbb{R}^n)} \le K_{GN} \|\nabla u\|^{\theta}_{L^2(\mathbb{R}^n)} \|u\|^{1-\theta}_{L^q(\mathbb{R}^n)}, \tag{1.1}$$

*Department of Mathematics, University of Pavia, and IMATI CNR, via Ferrata 1, 27100 Pavia, Italy. giuseppe.toscani@unipv.it

for $u \in W^{1,q}(\mathbb{R}^n)$. In (1.1) K_{GN} is a positive constant, $n > 2$ and $1 < q < p < 2^* = 2n/(n-2)$, or $n = 1, 2$ and $1 < q < p$. Last $\theta = [2n(1 - q/p)]/[2n - q(n-2)]$. Here we denoted

$$W^{1,q}(\mathbb{R}^n) = \{u \in L^q(\mathbb{R}^n) : \nabla u \in L^2(\mathbb{R}^n)\}.$$

and the L^2-norm of ∇u has been considered for simplicity, though in general, when $n > 2$ the Gagliardo–Nirenberg inequalities can be stated with the L^r-norm of ∇u, where $1 < r < n$.

The problem of finding the sharp constants and optimal functions for these inequalities has attracted many researchers in the past years [1, 2, 16, 20, 31]. Also, results on their stability with respect to optimizers has been studied in various aspects (cf. [13, 24] and the references therein).

A maybe not so well-known aspect of a special class of these inequalities, is their relationships with information theory and entropy-type inequalities [40, 46, 47].

In information theory, the main examples are the celebrated Shannon entropy power inequality, formulated by Shannon in his pioneering paper [41], and the Blachman–Stam inequality [10, 42]. Given a random vector X in $\mathbb{R}^n$, $n \geq 1$ with density $f(x)$, let

$$\mathcal{H}(X) = \mathcal{H}(f) = -\int_{\mathbb{R}^n} f(x) \log f(x)\, dx \tag{1.2}$$

denote its entropy functional (or Shannon entropy). Together with entropy, Shannon introduced the concept of entropy power, defined by

$$\mathcal{N}(X) = \mathcal{N}(f) = \exp\left(\frac{2}{n}\mathcal{H}(X)\right). \tag{1.3}$$

The entropy power is built to be in a suitable sense *linear* at Gaussian random vectors. To be precise, let $Z_\sigma = N(0, \sigma I_n)$ denote the n-dimensional Gaussian random vector having mean vector 0 and covariance matrix σI_n, where I_n is the identity matrix. Then, the entropy power of Z_σ is linear in terms of the variance, since $\mathcal{N}(Z_\sigma) = \sigma$. Shannon entropy power inequality, first rigorously proven by Stam [42] (cf. also [17, 28, 29, 39, 46] for other proofs and extensions) gives a lower bound on Shannon entropy power of the sum of independent random variables X, Y in $\mathbb{R}^n$ with densities

$$\mathcal{N}(X+Y) \geq \mathcal{N}(X) + \mathcal{N}(Y), \tag{1.4}$$

with equality if and only if X and Y are Gaussian random vectors with proportional covariance matrices. In other words, Shannon entropy power characterizes an extremal property of Gaussian functions with respect to convolutions.

Likewise, Blachman–Stam inequality is concerned with the behavior of the Fisher information with respect to convolutions. Historically, it was the key argument to prove Shannon entropy power inequality [10, 42]. Given the n-dimensional random vector X of probability density $f(x)$, let

$$\mathcal{I}(X) = \mathcal{I}(f) = \int_{\mathbb{R}^n} f(x) \left|\nabla \log f(x)\right|^2 \, dx = \int_{\{f>0\}} \frac{|\nabla f(x)|^2}{f(x)} \, dx. \tag{1.5}$$

define its Fisher measure of information. Analogously to the entropy power inequality (1.4), Blachman–Stam inequality furnishes a lower bound on the Fisher information of the sum of independent random variables X, Y in $\mathbb{R}^n$ with densities

$$\frac{1}{\mathcal{I}(X+Y)} \geq \frac{1}{\mathcal{I}(X)} + \frac{1}{\mathcal{I}(Y)}, \tag{1.6}$$

with equality if and only if X and Y are Gaussian random vectors with proportional covariance matrices. Hence, analogously to (1.4), inequality (1.6) characterizes a further extremal property of Gaussian functions with respect to convolutions. Note that, similarly to the entropy power, $1/\mathcal{I}(X)$ shares the same *linearity property* of the entropy power. Indeed $1/\mathcal{I}(X) = n\sigma$ when $X = Z_\sigma$.

In Shannon's theory of information, the proof of these inequalities has benefited from the close links between entropies and diffusion equations. These links are perfectly understood in the linear case, where, in addition to its paramount importance in physical applications, the linear heat equation is known to represent a profitable tool to obtain mathematical inequalities in sharp form [6, 7, 8, 9, 45, 46].

This original way of using heat equation dates back to the years between the late fifties to mid sixties, exactly at the same time in which Gagliardo [27] and Nirenberg [37] proved the interpolation inequalities that bear their name. The pioneering application of heat equation to the finding of analytic inequalities is due to Stam [42] in 1959 (cf. also Blachman [10]), who used the link between entropy and Fisher measure of information, obtained by deriving Shannon entropy along the solution to the heat equation, to find a rigorous proof of inequality (1.4). It is interesting to remark that the same link was used independently by Linnik [32] (his research appeared in the same year of Stam's work [42]) to get a new information-theoretic proof of the central limit theorem of probability theory.

Also, some years later, the heat equation has been used in the context of kinetic theory of rarefied gases by McKean [30] to investigate the large-time behavior of Kac caricature of a Maxwell gas. There, various monotonicity properties of the derivatives of Shannon entropy along the solution to the heat equation have been obtained.

These contributions made popular with the information community the deep link between Shannon entropy (1.2) and the solution of the heat equation posed in the whole space $\mathbb{R}^n$

$$\frac{\partial u(x,t)}{\partial t} = \Delta u(x,t), \tag{1.7}$$

as soon as the initial datum is assumed to be a probability density.

More recently [25, 40, 47], it has noticed that a similar link can be established between Rényi entropy of order p and the nonlinear diffusion of order p, posed in the whole space $\mathbb{R}^n$

$$\frac{\partial v(x,t)}{\partial t} = \Delta v^p(x,t), \tag{1.8}$$

still with the initial datum assumed to be a probability density, in the range $p > (n-2)/n$.

Given a random vector X in $\mathbb{R}^n$, $n \geq 1$ with density $f(x)$, and a positive constant p, Rényi entropy of order p of X is defined by [22]

$$\mathcal{H}_p(X) = \mathcal{H}_p(f) = \frac{1}{1-p} \log \left(\int_{\mathbb{R}^n} f^p(y)\, dy \right). \tag{1.9}$$

This concept of entropy was introduced by Rényi in [38] for a discrete probability measure to generalize the classical logarithmic entropy, by maintaining at the same time most of its properties. Indeed, Rényi entropy of order 1, defined as the limit as $p \to 1$ of $\mathcal{H}_p(f)$ coincides with Shannon entropy. Therefore, the standard (Shannon) entropy of a probability density [41] is included in the set of Rényi entropies, and it is identified with Rényi entropy of index $p = 1$.

The deep link between nonlinear diffusions and Rényi entropies gave a new light to a certain class of inequalities already present in the literature. Indeed, inequalities for Rényi entropies and generalized Fisher information measures were considered before [34, 35], without resorting to any connection with nonlinear diffusions.

While the derivation of sharp inequalities involving entropies, and the (eventual) characterization of the extremal densities, represents one of the main objectives of information theory [19, 22], often this point of view and the subsequent results do not spread automatically to other mathematical communities.

The aim of this note is to outline the close relationships between entropy type and theoretic inequalities. In particular, we will show that a certain class of Gagliardo–Nirenberg type interpolation inequalities can be viewed as extremal properties of Rényi measures of information. Consequently, their derivation can be obtained by resorting to arguments which are strongly based on their information-theoretic meaning. These results are rooted in many contributions from the field of information theory [17, 19, 18, 22, 34, 35, 49], mass transportation and nonlinear diffusion equations [11, 12, 14, 16, 18, 20]. Also, details on some of these results can be extracted from various recent papers of this author [15, 23, 25, 40, 44, 45, 46, 47].

2 Heat equation and related entropies

The goal of this Section is to enlighten, by means of a simple example, how the solution to the linear heat equation can play a role in recovering information-type inequalities in sharp form. To start with, let us briefly recall some elementary but basic result about the linear diffusion equation. Further details can be found in any introductory book on partial differential equations (cf. for example the classical treatise of Evans [26]).

Let $u(x,t)$ denote the solution to the initial value problem of equation (1.7), corresponding to an initial value $u_0(x)$ that is a probability density with finite second moment. The solution to the initial value problem for equation (1.7) is given by the convolution product of the initial density $u_0(x)$ with the Gaussian density M_{2t}, being

$$M_\sigma(x) = \frac{1}{(2\pi\sigma)^{n/2}} \exp\left\{-\frac{x^2}{2\sigma}\right\}. \tag{2.1}$$

the Gaussian density of the n-dimensional Gaussian random vector Z_σ of mean vector zero and covariance matrix σI (with variance $n\sigma$).

It is well-known (and in any case immediate to verify) that the solution to the heat equation is such that mass and momentum are preserved in time, so that

$$\rho(u(t)) = \int_{\mathbb{R}^n} u(x,t)\,dx = \int_{\mathbb{R}^n} u_0(x)\,dx = 1,$$

and

$$m(u(t)) = \int_{\mathbb{R}^n} x u(x,t)\,dx = \int_{\mathbb{R}^n} x u_0(x)\,dx = m(u_0).$$

Differently, the second moment

$$E(u(t)) = \int_{\mathbb{R}^n} |x|^2 u(x,t)\,dx$$

is growing linearly in time, at a rate given by

$$\frac{d}{dt} E(u(t)) = 2n\rho(u(t)) = 2n. \tag{2.2}$$

Note that the growth of the second moment depends on the initial value only through its mass density. This implies that, in correspondence to each initial datum of unit mass, independently of its shape, the rate of growth has the same value $2n$. However, a more refined estimate can be obtained by resorting to the forthcoming argument, which takes advantage of a particular rewriting of the Laplace operator.

Since the initial value $u_0(x)$ is a probability density, $u_0(x) \geq 0$, and the maximum principle insures that the solution $u(x,t)$ remains nonnegative for all subsequent times $t > 0$. Thus, we can write the diffusion equation (1.7) in the alternative form

$$\frac{\partial u(x,t)}{\partial t} = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} \left(u(x,t) \frac{\partial}{\partial x_i} \log u(x,t) \right). \tag{2.3}$$

For any given time $t > 0$, integration by parts gives

$$\begin{aligned} \frac{d}{dt} E(u(t)) =& \frac{d}{dt} \int_{\mathbb{R}^n} |x|^2 u(x,t)\,dx = \int_{\mathbb{R}^n} |x|^2 \frac{\partial u(x,t)}{\partial t}\,dx = \\ & \int_{\mathbb{R}^n} |x|^2 \sum_{i=1}^{n} \frac{\partial}{\partial x_i} \left(u(x,t) \frac{\partial}{\partial x_i} \log u(x,t) \right) dx = \\ & -2 \int_{\mathbb{R}^n} \sum_{i=1}^{n} x_i u(x,t) \frac{\partial}{\partial x_i} \log u(x,t)\,dx. \end{aligned} \tag{2.4}$$

By Cauchy–Schwarz inequality

$$\left| \int_{\mathbb{R}^n} \sum_{i=1}^{n} x_i \frac{\partial}{\partial x_i} \log u(x,t)\, u(x,t)\, dx \right| \le E(u(t))^{1/2} \mathcal{I}(u(t))^{1/2}, \tag{2.5}$$

being $\mathcal{I}(u(t))$ the Fisher information of the solution (cf. definition (1.5)). We remark that, as proven in [33], Lemma 2.1, $\mathcal{I}(u(t))$ is bounded as soon as $t > 0$. Inequality (2.5) then implies that the growth of the square root of the second moment of the solution to the heat equation can not exceed the square root of the Fisher information of the solution itself. Indeed, substituting the right-hand side of (2.5) into the last line of (2.4) we get

$$\frac{d}{dt}\sqrt{E(u(t))} \le \sqrt{\mathcal{I}(u(t))}. \tag{2.6}$$

Moreover, using (2.2) into (2.6) we conclude with the information-type inequality [22]

$$E(u(t))\mathcal{I}(u(t)) \ge n^2. \tag{2.7}$$

Note that both inequality (2.5) and the differential inequality (2.6) become equalities when evaluated in correspondence to the Gaussian M_{2t}, $t > 0$, fundamental solution of the heat equation (1.7).

Indeed, for the Gaussian function $M_\sigma(x)$, it holds

$$-\int_{\mathbb{R}^n} \sum_{i=1}^{n} x_i M_\sigma(x) \frac{\partial}{\partial x_i} \log M_\sigma(x)\, dx = n,$$

while

$$\mathcal{I}(M_\sigma) = \frac{n}{\sigma}.$$

Let us draw some conclusions from the previous example. We studied the evolution of a time dependent *functional* of the solution to the heat equation (the second moment) which grows linearly with respect to time. In the case of the linear diffusion, this linear growth is common to all solutions.

Since the growth is linear, the time derivative of the functional is constant. In particular, this is true if we consider the evolution of the second moment of the fundamental solution M_{2t}. Note however that time appears in the fundamental solution as a dilation parameter, where, as usual, for any given density $f(x)$ and positive constant a, we define the *dilation* of f by a, as the mass-preserving scaling

$$f(x) \to f_a(x) = a^n f(ax). \tag{2.8}$$

Consequently, the time derivative of the functional, which is constant in time, is invariant with respect to dilations (scale invariant). Hence, we can take the limit as $t \to 0^+$ of the left–hand side of inequality (2.7) to obtain, for any probability density $f(x)$, $x \in \mathbb{R}^n$, with bounded Fisher information the scale invariant inequality

$$E(f)\mathcal{I}(f) \geq n^2. \tag{2.9}$$

Note that the constant n^2 in inequality (2.9) is sharp on the set of probability densities. This value is reached in correspondence to Gaussian densities, which are the unique minimizers. Moreover, for any given initial value $u_0(x)$ which is a probability density, the Gaussian density is the unique initial value for which equality in (2.6) holds.

A further example will help to understand the remarkable potential of this idea. Let us study the time evolution of Shannon entropy, as defined in (1.2), of the solution to the heat equation. Using the heat equation written in the form (2.3) we easily conclude with DeBruijn's identity

$$\frac{d}{dt}\mathcal{H}(u(t)) = \mathcal{I}(u(t)), \quad t > 0, \tag{2.10}$$

that links Shannon entropy with Fisher measure of information via the heat equation. Now, consider that Shannon entropy of the fundamental solution to the heat equation has the value

$$\mathcal{H}(M_{2t}) = \frac{n}{2}\log(4\pi e t). \tag{2.11}$$

Hence, it grows logarithmically with time. To restore the linear growth we need to consider the quantity $\exp\{2\mathcal{H}(f)/n\}$, namely Shannon entropy power of a probability density defined in (1.3). This clarifies (from the point of view of the linear diffusion equation) why the entropy power is the right quantity to study to get inequalities.

As in the previous example, let us compute the subsequent derivatives of Shannon entropy power. These computations lead to a well-known result in information theory, known as the *concavity of entropy power*. If $u(x,t)$ is a solution to the heat equation (1.7), corresponding to an initial datum $u_0(x)$ that is a probability density, then its entropy power is a concave function of time

$$\frac{d^2}{dt^2}\mathcal{N}(u(t)) \leq 0. \tag{2.12}$$

Moreover, equality in (2.12) holds if and only if $u(x,t)$ coincides with the Gaussian density of variance $2t$, namely the fundamental solution to the heat equation. Inequality (2.12) has been first obtained by Costa [17]. Simplified proofs were subsequently done by Dembo [21] by means of a direct application of Blachman–Stam inequality (1.6), and Villani [49], who made use of an argument introduced by McKean [30] in his paper on Kac caricature of a Maxwell gas.

Maybe the most important consequence of (2.12) is the sharp inequality which can be extracted from the first derivative of the entropy power [44]. Indeed, analogously to the previous example, the first derivative is both scale invariant, and nonincreasing in consequence of the concavity property. Since

$$\frac{d}{dt}\mathcal{N}(u(t)) = \frac{2}{n}\mathcal{N}(u(t))\mathcal{I}(u(t)), \tag{2.13}$$

invariance under dilation coupled with concavity allow to reach the lower bound by taking the limit as $t \to \infty$ of the right-hand side of equation (2.13). This implies the so-called *isoperimetric inequality for entropies* [22, 44]. For any probability density $f(x)$ in $\mathbb{R}^n$, this scale invariant inequality asserts that

$$\mathcal{N}(f)\,\mathcal{I}(f) \geq 2\pi e n. \tag{2.14}$$

As for (2.7), the constant $2\pi e n$ is sharp on the set of probability densities. This value is reached in correspondence to Gaussian densities, which are the unique minimizers. Moreover, for any given initial value $u_0(x)$ which is a probability density, the Gaussian density is the unique initial value for which equality in (2.12) holds.

It is interesting to remark that from the isoperimetric inequality for entropies (2.14) one can obtain, among other consequences, the logarithmic Sobolev inequality in scale invariant form, and Nash's inequality [36] with an asymptotically sharp constant [44]. In particular, Nash's interpolation inequality reads

$$\left(\int_{\mathbb{R}^n} g^2(x)\,dx\right)^{1+2/n} \leq \frac{2}{\pi e n}\left(\int_{\mathbb{R}^n} |g(x)\,dv\right)^{4/n} \int_{\mathbb{R}^n} |\nabla g(x)|^2\,dx. \tag{2.15}$$

3 From Rényi's entropies to Gagliardo–Nirenberg inequalities

The arguments of Section 2, enlightening the role of the linear heat equation in recovering sharp inequalities, can be generalized to the nonlinear diffusion equations (1.8) and relative entropies.

Similarly to Section 2, our analysis will be restricted to initial data which are probability densities with finite variance. This allows to include both the case $p > 1$, usually known with the name of porous medium equation, and a limited range of exponents when $p < 1$, the fast diffusion equation. In dimension $n \geq 1$, the second moment of the solution to equation (1.8) is bounded if $p > \bar{p}$ with $\bar{p} = n/(n+2)$. The particular subinterval of p is motivated by the existence of a precise solution, found by Zel'dovich, Kompaneets and Barenblatt in the fifties (briefly called here Barenblatt solution) [4, 5, 50], which serves as a model for the asymptotic behavior of a wide class of solutions with finite second moment. In the case $p > 1$ (see [11] for $p < 1$) the Barenblatt (also called self-similar or generalized Gaussian solution) departing from $x = 0$ takes the self-similar form

$$M_p(x,t) := \frac{1}{t^{n/\mu}} \tilde{M}_p\left(\frac{x}{t^{1/\mu}}\right), \tag{3.1}$$

where

$$\mu = 2 + n(p-1)$$

and $\tilde{M}_p(x)$ is the time-independent function

$$\tilde{M}_p(x) = \left(C - \lambda\, |x|^2\right)_+^{\frac{1}{p-1}}. \tag{3.2}$$

In (3.2) $(s)_+ = \max\{s, 0\}$, $\lambda = \frac{1}{2\mu}\,\frac{p-1}{p}$, and the constant C is chosen to fix the mass of the source-type Barenblatt solution equal to one.

Since we are interested in presenting the information-theoretic aspect of inequalities, in what follows, we will often proceed formally. However, at the price of an increasing number of technical details, the results can be fully justified on the basis of classical results of existence, uniqueness and regularity of the solution to the initial value problem [11, 40, 48]. In analogy with the heat equation, the solution to the nonlinear diffusion (1.8) is such that mass and momentum are preserved in time

$$\rho(v(t)) = \int_{\mathbb{R}^n} v(x,t)\, dx = \int_{\mathbb{R}^n} v_0(x)\, dx = 1,$$

and

$$m(v(t)) = \int_{\mathbb{R}^n} x v(x,t)\, dx = \int_{\mathbb{R}^n} x v_0(x)\, dx = m(v_0).$$

Moreover, while the second moment $E(v(t))$ of equation (1.8) increases in time from $E_0 = E(v_0)$, differently from the heat equation its evolution is now given by the nonlinear law

$$\frac{dE(v(t))}{dt} = 2n \int_{\mathbb{R}^n} v^p(x,t)\, dx \geq 0, \tag{3.3}$$

which, since $p \neq 1$, it is not explicitly computable.

Note that in the nonlinear case the growth of the second moment varies with time, and depends on the solution itself. Nevertheless, one can try to apply to equation (1.8) the same strategy applied to the linear heat equation. Similarly to the case $p = 1$, we can write the nonlinear diffusion equation in the alternative form

$$\frac{\partial v(x,t)}{\partial t} = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} \left(p\, v(x,t) \frac{\partial}{\partial x_i} \frac{v^{p-1}(x,t)}{p-1}(x,t) \right). \tag{3.4}$$

For any given time $t > 0$, integration by parts gives

$$\begin{aligned} \frac{d}{dt}E(v(t)) =& \frac{d}{dt} \int_{\mathbb{R}^n} |x|^2 v(x,t)\, dx = \int_{\mathbb{R}^n} |x|^2 \frac{\partial v(x,t)}{\partial t}\, dx = \\ & \int_{\mathbb{R}^n} |x|^2 \sum_{i=1}^{n} \frac{\partial}{\partial x_i} \left(p\, v(x,t) \frac{\partial}{\partial x_i} \frac{v^{p-1}(x,t)}{p-1}(x,t) \right) dx = \\ & -2\, p \int_{\mathbb{R}^n} \sum_{i=1}^{n} x_i v(x,t)\, \frac{\partial}{\partial x_i} \frac{v^{p-1}(x,t)}{p-1}(x,t)\, dx. \end{aligned} \tag{3.5}$$

By Cauchy–Schwarz inequality

$$\begin{aligned} & \left| \int_{\mathbb{R}^n} \sum_{i=1}^{n} x_i \frac{\partial}{\partial x_i}\, v(x,t) \frac{v^{p-1}(x,t)}{p-1}(x,t)\, dx \right| \leq \\ & E(v(t))^{1/2} \left(\int_{\mathbb{R}^n} v(x,t) \left| \frac{\nabla v^{p-1}(x,t)}{p-1} \right|^2 dx \right)^{1/2} = \\ & \frac{1}{p} E(v(t))^{1/2} \left(\int_{\mathbb{R}^n} \frac{|\nabla v^p(x,t)|^2}{v(x,t)}\, dx \right)^{1/2}. \end{aligned} \tag{3.6}$$

For a given n-dimensional random vector X of probability density $f(x)$, let us define the Fisher information of order p as [40]

$$\mathcal{I}_p(X) = \mathcal{I}_p(f) := \frac{1}{\int_{\mathbb{R}^n} f^p(x)\, dx} \int_{\{f>0\}} \frac{|\nabla f^p(x)|^2}{f(x)}\, dx. \tag{3.7}$$

We remark that, as $p \to 1$ the Fisher information of order p tends towards the classical Fisher information defined by (1.5). Then, from equality (3.3) we obtain the bound

$$E(v(t))I_p(v(t)) \geq n^2 \int_{\mathbb{R}^n} v^p(x,t)\,dx. \tag{3.8}$$

By taking the limit as $t \to 0^+$ of both sides of inequality (3.8), it follows that any probability density $f(x)$, $x \in \mathbb{R}^n$, with bounded variance satisfies the inequality

$$E(f)I_p(f) \geq n^2 \int_{\mathbb{R}^n} f^p(x)\,dx. \tag{3.9}$$

As before, equality is reached in (3.9) if and only if $f = \tilde{M}_p$, the Barenblatt profile of order p. Note that, at difference with the linear case, when $p \neq 1$, the derivative of the second moment can be directly related to Rényi entropy. Indeed, by (1.9)

$$\int_{\mathbb{R}^n} f^p(x)\,dx = \exp\left\{(1-p)\mathcal{H}_p(f)\right\}. \tag{3.10}$$

Hence, we can rewrite inequality (3.8) as an information-type inequality

$$E(f)I_p(f) \geq n^2 \exp\left\{(1-p)\mathcal{H}_p(f)\right\}, \tag{3.11}$$

which contains inequality (2.7) as $p \to 1$.

The definition of the p-Fisher information is motivated by the following relationship. Let us consider the evolution in time of Rényi entropy of order p, defined in (1.9), along the solution of the nonlinear diffusion equation (1.8). Integration by parts then yields, for $t > 0$

$$\frac{d}{dt}\mathcal{H}_p(v(\cdot,t)) = \mathcal{I}_p(v(\cdot,t)). \tag{3.12}$$

When $p \to 1$, identity (3.12) reduces to DeBruijn's identity (2.10) which connects Shannon entropy functional with the Fisher information.

Before to proceed, let us recall some interesting information-theoretic consequences of inequality (3.8) (alternatively (3.11)). Among other properties, Rényi entropy (1.9) behaves as Shannon entropy (1.2) with respect to dilations of the probability density. Indeed, for any $p \geq 0$ it holds

$$\mathcal{H}_p(f_a) = \mathcal{H}_p(f) - n\log a. \tag{3.13}$$

This property implies that, for any given probability density function f in $\mathbb{R}^n$ such that both the second moment and Rényi entropy are bounded, the functional

$$\Lambda(f) = \mathcal{H}_p(f) - \frac{n}{2} \log E(f) \tag{3.14}$$

is invariant with respect to dilations. Let us consider the time variation of $\Lambda(v(t))$, where $v(x,t)$ is the solution to the nonlinear diffusion (1.8) corresponding to an initial value $u_0(x)$ which is a probability density with finite second moment. Clearly, thanks to inequality (3.8) it holds

$$\frac{d}{dt}\Lambda(v(t)) = \mathcal{I}_p(v(t)) - n^2 \frac{\int_{\mathbb{R}^n} v^p(x,t)\,dx}{E(v(t))} \geq 0. \tag{3.15}$$

Thus, $\Lambda(v(t))$ is increasing in time, which implies, grace to dilation invariance, that it will converge, as time $t \to \infty$ to the value obtained in correspondence to the Barenblatt profile [47]. Hence, given a probability density function $f(x)$, $x \in \mathbb{R}^n$, such that both the second moment and Rényi entropy are bounded, it holds

$$\mathcal{H}_p(f) - \frac{n}{2} \log E(f) \leq \mathcal{H}_p(\tilde{M}) - \frac{n}{2} \log E(\tilde{M}_p). \tag{3.16}$$

The same inequality has been proven in [18, 34, 35] by different methods.

Let $\mathcal{N}_p(f)$ denote the entropy power of f associated to Rényi's entropy of order p

$$\mathcal{N}_p(f) = \exp\left\{ \left(\frac{2}{n} + p - 1 \right) \mathcal{H}_p(f) \right\}. \tag{3.17}$$

Then, if $p > n/(n+2)$, we rewrite (3.16) as the information type inequality

$$\frac{\mathcal{N}_p(f)}{E(f)^{1+n(p-1)/2}} \leq \frac{\mathcal{N}_p(B_{p,\sigma})}{E(B_{p,\sigma})^{1+n(p-1)/2}}. \tag{3.18}$$

We remark that the definition (3.17) of p-Rényi entropy power, proposed recently in [40], coincides with the classical definition of Shannon entropy power [41], valid when $p = 1$, since it has been constructed to be linear in time in correspondence to the Barenblatt solution of the nonlinear diffusion equation (1.8). Indeed, owing to definition (3.17), it is a simple exercise to show that

$$\mathcal{N}_p(M_p(t)) = \mathcal{N}_p(\tilde{M}_p) \cdot t. \tag{3.19}$$

Definition (3.17) requires $p > (n-2)/n$, in which case $2/n + p - 1 > 0$. The range of the parameter p for which we can introduce our notion of Rényi entropy power, coincides with the range for which there is mass conservation for the solution of (1.8) [11]. This range includes the cases in which the Barenblatt has bounded second moment, since $(n-2)/n < n/(n+2)$.

Definition (3.17) allows to extend the concavity property of Costa [17], relative to the linear diffusion equation, to Rényi entropy power of the solution to the nonlinear diffusion equation (1.8), as soon as the second moment of the Barenblatt solution is bounded [40]. The precise result is the following. Let $p > n/(n+2)$ and let $u(\cdot,t)$ be probability densities in $\mathbb{R}^n$ solving the initial value problem for equation (1.8) for $t > 0$. Then the p-th Rényi entropy power defined by (3.17) has the *concavity property* so that

$$\frac{d^2}{dt^2}\mathcal{N}_p(v(\cdot,t)) \le 0. \tag{3.20}$$

Moreover, equality in (3.20) is achieved if and only if $v(t)$ is the Barenblatt solution (3.1). Like in Shannon's case, inequality (3.20) lieds to sharp isoperimetric inequalities. Under the allowed conditions on the parameter p, the first (nonincreasing) derivative of p-th Rényi entropy power along the solution to the nonlinear diffusion gives the scale invariant *isoperimetric inequality*

$$\mathcal{N}_p(f)\,\mathcal{I}_p(f) \ge \mathcal{N}_p(\tilde{M}_p)\,\mathcal{I}_p(\tilde{M}_p) = \gamma_{n,p}. \tag{3.21}$$

Indeed, it is immediate to show that the product in (3.21) is invariant under dilation, which allows to reckon explicitly the value of the constants $\gamma_{n,p}$ by using the same argument of Section 2 [40]. Clearly, the limit value is obtained in correspondence to the Barenblatt profile (3.2).

Inequality (3.21) can be rewritten in a form more suitable to functional analysis. Indeed, when $f(x)$ be a probability density in $\mathbb{R}^n$, and $p > n/(n+2)$, the isoperimetric inequality (3.21) takes the form

$$\int_{\mathbb{R}^n} \frac{|\nabla f^p(x)|^2}{f(x)}\,dx \ge \gamma_{n,p}\left(\int_{\mathbb{R}^n} f^p(x)\,dx\right)^{\frac{2+2n(p-1)}{n(p-1)}}. \tag{3.22}$$

By setting into (3.22)

$$f^{p-1/2} = \frac{u}{\|u\|_{L^{2q}(\mathbb{R}^n)}}, \quad q = \frac{1}{2p-1},$$

one obtains that f is a probability density in $\mathbb{R}^n$, and

$$\int_{\mathbb{R}^n} \frac{|\nabla f^p(x)|^2}{f(x)}\,dx = 4p^2q^2 \int_{\mathbb{R}^n} |\nabla u(x)|^2\,dx \left(\int_{\mathbb{R}^n} u^{2q}\,dx\right)^{-1/q}.$$

Hence one realizes that inequality (3.22) is equivalent to one of the two following Gagliardo–Nirenberg inequalities [25]. If $(n-1)/n \le p < 1$ then (3.22) is equivalent to

$$\|u\|_{L^{2q}(\mathbb{R}^n)} \le K_{GN} \|\nabla u\|^{\theta}_{L^2(\mathbb{R}^n)} \|u\|^{1-\theta}_{L^{q+1}(\mathbb{R}^n)}, \tag{3.23}$$

where

$$1 < q \le \frac{n}{n-2}, \quad \theta = \frac{n}{q} \frac{q-1}{n+2-q(n-2)}.$$

If $p > 1$, then (3.22) is equivalent to

$$\|u\|_{L^{q+1}(\mathbb{R}^n)} \le K_{GN} \|\nabla u\|^{\theta}_{L^2(\mathbb{R}^n)} \|u\|^{1-\theta}_{L^{2q}(\mathbb{R}^n)}, \tag{3.24}$$

where

$$0 < q < 1, \quad \theta = \frac{n}{q} \frac{q-1}{n+2-q(n-2)}.$$

In particular, *Sobolev* inequality [3, 43] is obtained when $\theta = 1$, $p = 1/n$ and $2q = n/(n-2)$. Hence, Sobolev inequality with the sharp constant is a consequence of the concavity of Rényi entropy power of parameter $p = (n-1)/n$, when $n > 2$, and represents the threshold case for the validity of the method. These Gagliardo-Nirenberg type inequalities with sharp constants, have been first obtained by Del Pino and Dolbeault [20], and later on by Cordero-Erausquin, Nazaret, and Villani, [16] with different methods, without noticing their connection with classical concepts of information theory. However, in the one-dimensional case, these connections have been outlined in [34], where a more general class of generalized Fisher-type measures of information have been considered, together with the corresponding isoperimetric inequalities.

4 Conclusions

In this note we presented the principal connections between information-theoretic inequalities for entropies and interpolation inequalities of Gagliardo–Nirenberg type. It is interesting to remark that the proof of most information-type inequalities takes advantage of the time-evolution of entropies along the solutions to diffusion equations, which are known to converge towards self-similar ones. The main example is furnished by the result of concavity presented in Section 3, where the classical sharp Sobolev inequality is derived as a particular case of the concavity property of Rényi entropy power of order $1 - 1/n$. Like in the linear case, the knowledge of isoperimetric inequalities for entropies is a powerful tool that, among other applications, allows to obtain sharp convergence results on the large-time behavior of the solution to nonlinear diffusion equations [15].

Acknowledgement

This work has been written within the activities of GNFM group of INdAM (National Institute of High Mathematics), and partially supported by MIUR project "Optimal mass transportation, geometrical and functional inequalities with applications". The research was partially supported by the Italian Ministry of Education, University and Research (MIUR): Dipartimenti di Eccellenza Program (2018–2022) - Dept. of Mathematics "F. Casorati", University of Pavia.

I wish to thank the editors of this volume for inviting me to contribute to the memory of the notable figure of Emilio Gagliardo, who was an esteemed colleague in my Department for more than twenty years, starting in November 1975.

References

[1] M. Agueh, N. Ghoussoub, X. Kang, Geometric inequalities via a general comparison principle for interacting gases. *Geom. Funct. Anal.*, **14** (2004) 215–244

[2] M. Agueh, Gagliardo–Nirenberg inequalities involving the gradient L^2-norm. *C. R. Acad. Sci. Paris, Ser. I*, **346** (2008) 757–762

[3] T. Aubin, Problémes isopérimétriques et espaces de Sobolev. *J. Differential Geometry*, **11** (1976)

[4] G.I. Barenblatt, On some unsteady motions of a liquid and gas in a porous medium. *Akad. Nauk SSSR. Prikl. Mat. Meh.*, **16** (1952) 67–78.

[5] G.I. Barenblatt, *Scaling, self-similarity and intermediate asymptotics.* Cambridge Univ. Press, Cambridge 1996.

[6] J. Bennett, Heat-flow monotonicity related to some inequalities in euclidean analysis. *Contemporary Mathematics*, **505** (2010), 85–96.

[7] J. Bennett and N. Bez, Closure properties of solutions to heat inequalities. *J. Geom. Anal.*, **19** (2009), 584–600.

[8] J. Bennett, N. Bez and A. Carbery, Heat-flow monotonicity related to the Hausdorff–Young inequality, *Bull. London Math. Soc.* **41** (6) (2009), 971–979.

[9] J. Bennett, A. Carbery, M. Christ and T. Tao, The Brascamp–Lieb inequalities: finiteness, structure and extremals. *Geometric And Functional Analysis*, **17** (2007), 1343–1415.

[10] N.M. Blachman, The convolution inequality for entropy powers. *IEEE Trans. Inform. Theory*, **11** (1965) 267–271.

[11] A. Blanchet, M. Bonforte, J. Dolbeault, G. Grillo, and J. L Vázquez, Asymptotics of the fast diffusion equation via entropy estimates. *Arch. Ration. Mech. Anal.*, **191** (2009) 347–385.

[12] M. Bonforte, J. Dolbeault, G. Grillo, and J.-L. Vázquez, Sharp rates of decay of solutions to the nonlinear fast diffusion equation via functional inequalities. *Proc. Natl. Acad. Sci. USA*, **107** (2010) 16459–16464.

[13] E.A. Carlen, and A. Figalli, Stability for a GNS inequality and the Log-HLS inequality, with application to the critical mass Keller–Segel equation. *Duke Math. J.*, **162** (3) (2013) 579–625.

[14] J. A. Carrillo, and G. Toscani, Asymptotic L^1-decay of solutions of the porous medium equation to self-similarity. *Indiana Univ. Math. J.*, **49** (2000) 113–141.

[15] J.A. Carrillo, and G. Toscani, Rényi entropy and improved rate of convergence to self-similarity for nonlinear diffusion equations. *Nonlinearity*, **27** (2014) 3159–3177.

[16] D. Cordero-Erausquin, B. Nazaret, and C. Villani, A mass–transportation approach to sharp Sobolev and Gagliardo–Nirenberg inequalities. *Advances in Mathematics*, **182** (2004) 307–332.

[17] M. Costa, A new entropy power inequality. *IEEE Trans. Inform. Theory*, **31** (1985) 751–760.

[18] J. Costa, A. Hero, and C. Vignat, On solutions to multivariate maximum alpha-entropy problems. *Lecture Notes in Computer Science*, **2683**, no. EMMCVPR 2003, Lisbon, 7-9 July 2003, (2003) 211–228.

[19] T.M. Cover, and J.A. Thomas, *Elements of information theory.* Wiley & Sons, New York, 1992.

[20] M. Del Pino, and J. Dolbeault, Best constants for Gagliardo–Nirenberg inequalities and application to nonlinear diffusions. *J. Math. Pures Appl.*, **81** (2002) 847–875.

[21] A. Dembo, A simple proof of the concavity of the entropy power with respect to the variance of additive normal noise. *IEEE Trans. Inform. Theory*, **35** (1989) 887–888.

[22] A. Dembo, T. Cover, and J. Thomas, Information theoretic inequalities. *IEEE Trans. Inform. Theory*, **37** (1991) 1501–1518.

[23] J. Dolbeault, and G. Toscani, Best matching Barenblatt profiles are delayed. *J. Phys. A: Math. Theor.*, **48** (2015) 065206.

[24] J. Dolbeault, and G. Toscani, Stability results for logarithmic Sobolev and Gagliardo–Nirenberg inequalities. *Int. Math. Res. Notices*, rnv 131 (2015).

[25] J. Dolbeault, and G. Toscani, Nonlinear diffusions: extremal properties of Barenblatt profiles, best matching and delays. *Nonlinear Anal. Series A*, **138** (2016) 31–43.

[26] L.C. Evans, *Partial differential equations.* Graduate Studies in Mathematics 19. AMS, Providense, RI 2010.

[27] E. Gagliardo, Proprietà di alcune classi di funzioni in più variabili. *Ricerche Mat.*, **7** (1958) 102–137.

[28] D. Guo, S. Shamai, and S. Verdú, Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Inf. Theory*, **51**, (4) (2005) 1261–1282.

[29] D. Guo, S. Shamai, and S. Verdú, A simple proof of the entropy-power inequality . *IEEE Trans. Inf. Theory*, **52**, (5) (2006) 2165–2166.

[30] H.P. McKean Jr., Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas. *Arch. Rat. Mech. Anal.*, **21** (1966) 343–367.

[31] H.A. Levine, An estimate for the best constant in a Sobolev inequality involving three integral norms. *Ann. Mat. Pura Appl.*, **124** (1980) 181–197.

[32] Yu.V. Linnik, An information-theoretic proof of the central limit theorem with the Lindeberg condition. *Theory Probab. Appl.*, **4** (1959) 288-299.

[33] P.L.Lions, and G.Toscani, A sthrenghtened central limit theorem for smooth densities. *J. Funct. Anal.*, **128** (1995) 148–167.

[34] E. Lutwak, D. Yang, and G. Zhang, Cramér-Rao and moment-entropy inequalities for Rényi entropy and generalized Fisher information. *IEEE Trans. Inform. Theory*, **51** (2005) 473–478.

[35] E. Lutwak, D. Yang, and G. Zhang, Moment-entropy inequalities for a random vector. *IEEE Trans. Inform. Theory*, **53** (2007) 1603–1607.

[36] J.F. Nash, Continuity of solutions of parabolic and elliptic equations. *Amer. J. Math.*, **80** (1958) 931–954.

[37] L. Nirenberg, On elliptic partial differential equations. *Ann. Scuola Norm. Pisa*, **13** (1959) 116–162.

[38] A. Rényi, On measures of entropy and information. *Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1*, 547–561. University of California Press, Berkeley 1961.

[39] O. Rioul, Information theoretic proofs of entropy power inequalities. *IEEE Trans. Inform. Theory*, **57** (1) (2011) 33–55.

[40] G. Savaré, and G. Toscani, The concavity of Rényi entropy power. *IEEE Trans. Inform. Theory*, **60** (2014) 2687–2693.

[41] C.E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.*, **27** Jul.(1948), 379–423; Oct. (1948) 623–656.

[42] A.J. Stam, Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Contr.*, **2** (1959) 101–112.

[43] G. Talenti, Best constant in Sobolev inequality. *Ann. Mat. Pura Appl.*, **110** (1976) 353–372.

[44] G. Toscani, An information-theoretic proof of Nash's inequality. *Rend. Lincei Mat. Appl.*, **24** (2013) 83–93.

[45] G. Toscani, Lyapunov functionals for the heat equation and sharp inequalities. *Atti Acc. Peloritana Pericolanti, Classe Sc. Fis. Mat. e Nat.*, **91** (2013) 1–10.

[46] G. Toscani, Heat equation and convolution inequalities. *Milan J. Math.*, **82** (2014) 183–212.

[47] G. Toscani, Rényi entropies and nonlinear diffusion equations. *Acta. Appl. Math.*, **132** (2014) 595–604.

[48] J.L. Vazquez, *The porous medium equation: Mathematical theory.* Oxford University Press, Oxford 2007.

[49] C. Villani, A short proof of the concavity of entropy power. *IEEE Trans. Inform. Theory*, **46** (2000) 1695–1696.

[50] Ya.B. Zeldovich, and A.S. Kompaneetz, Towards a theory of heat conduction with thermal conductivity depending on the temperature. Collection of Papers Dedicated to 70th Birthday of Academician A. F. Ioffe, *Izd. Akad. Nauk SSSR*, Moscow, (1950) 61Ũ–71.